Annex to:

EFSA NDA Panel, 2022. Scientific Opinion on the Tolerable Upper Intake Level for dietary sugars. EFSA Journal 2022;20(2):7074. doi:10.2903/j.efsa.2022.7074

Annex I – Customised forms for the appraisal of the risk of bias of human studies

Table of Contents

Annex I – Customised forms for the appraisal of the risk of bias of human studies 1
1. Introduction.....3
2. Intervention studies on metabolic diseases and dental caries.....3
3. Observational studies on metabolic diseases including pregnancy endpoints and dental caries 10
Glossary, abbreviations, and acronyms 16
References.....16

1. Introduction

The appraisal tool for human intervention and observational studies has been adapted from the NTP Risk of Bias (RoB) tool.[1]

The answer format for the RoB questions is as follows:

++	**Definitely Low** risk of bias	There is direct evidence of low risk-of-bias practices (May include specific examples of relevant low risk-of-bias practices)
+	**Probably Low** risk of bias	There is indirect evidence of low risk-of bias practices OR it is deemed that deviations of low risk-of bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias
-/NR	**Probably High** risk of bias	There is indirect evidence of high risk-of-bias practices OR there is insufficient information (e.g. not reported or "NR") provided about relevant risk-of bias practices
--	**Definitely High** risk of bias	There is direct evidence of high risk-of-bias practices (May include specific examples of relevant high risk-of-bias practices)

2. Intervention studies on metabolic diseases and dental caries

To assess the risk of bias (RoB) of intervention studies on metabolic diseases, endpoint variables were grouped into clusters considered to be potentially affected by the same type of bias, as follows:

a) Body weight, BMI, waist circumference and related endpoints
b) Body fat, ectopic fat deposition (muscle, liver, abdominal fat)
c) Blood lipids (total cholesterol, LDL-cholesterol, HDL-cholesterol and derived indices; fasting triglycerides), fasting glucose and insulin and derived indices, uric acid
d) Blood pressure
e) Measures of insulin sensitivity obtained either in steady-state conditions (during an euglycemic hyperinsulinemic clamp) or in non-steady state conditions (e.g. during an intravenous glucose with frequent sampling/minimal model assessment (IVGTT)) and indices of glucose tolerance derived from the oral glucose tolerance test (OGTT)
f) Measures of blood glucose control (fructosamine, glycated haemoglobin, glycated albumin)

The customised appraisal form for the appraisal of human intervention studies can be found in Table I1.

[1] Available at: https://ntp.niehs.nih.gov/whatwestudy/assessments/noncancer/riskbias/index.html

Table I1. Customised form for the appraisal of human intervention studies

[++: definitely low risk of bias, +: probably low risk of bias, -/NR: probably high risk of bias, --: definitely high risk of bias].

Question	Rating	Explanation for expert judgement
1. Was administered dose or exposure level adequately randomized? **Key question**	++	There is direct evidence that subjects were allocated to any study group (or intervention sequence for cross-over studies) including controls using a method with a random component. Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, shuffling cards or envelopes, throwing dice, or drawing of lots. Restricted randomization (e.g., blocked randomization) to ensure particular allocation ratios will be considered low risk of bias. Similarly, stratified randomization and minimization approaches that attempt to minimize imbalance between groups on important prognostic factors (e.g., body weight) will be considered acceptable.
	+	There is indirect evidence that subjects were allocated to study groups (or intervention sequence for cross-over studies) using a method with a random component (i.e., authors state that allocation was random, without description of the method used) **OR** it is deemed that allocation without a clearly random component during the study would not appreciably bias results (e.g. cross-over studies with no or unlikely carry-over effects)
	-	There is indirect evidence that subjects were allocated to study groups using a method with a non-random component ***NOTE**: Non-random allocation methods may be systematic but have the potential to allow participants or researchers to anticipate the allocation to study groups. Such "quasi-random" methods include alternation, assignment based on date of birth, case record number, or date of presentation to study.*
	NR	There is insufficient information provided about how subjects were allocated to study groups (or intervention sequence for cross-over studies)
	--	There is direct evidence that subjects were allocated to study groups (or intervention sequence for cross-over studies) using a non-random method including judgment of the clinician, preference of the participant, the results of a laboratory test or a series of tests, or availability of the intervention.
2. Was allocation to study groups adequately concealed? ***NOTE:** Allocation concealment and blinding are often confused.* Allocation concealment *involves not disclosing to patients and those involved in recruiting trial*	++	There is direct evidence that at the time of recruitment the research personnel and subjects did not know what study group subjects were allocated to, and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable. Acceptable methods used to ensure allocation concealment include central allocation (including telephone, web-based and pharmacy-controlled randomization); sequentially numbered drug containers of identical appearance; sequentially numbered, opaque, sealed envelopes; or equivalent methods
	+	There is indirect evidence that the research personnel and subjects did not know what study group subjects were allocated to and it is unlikely that they could have broken the blinding of allocation until after recruitment was complete and irrevocable **OR**

participants the allocation sequence before random allocation *occurs. The allocation sequence is the order in which participants are to be allocated to treatment.* Blinding *involves not disclosing to patients and outcome assessors the treatment allocations* after random allocation.		It is deemed that lack of adequate allocation concealment would not appreciably bias results (e.g. cross-over studies where all subjects receive all the study treatments)
	-	There is indirect evidence that at the time of recruitment it was possible for the research personnel and subjects to know what study group subjects were allocated to (or treatment sequence for cross-over studies), or it is likely that they could have broken the blinding of allocation before recruitment was complete and irrevocable ***NOTE:*** *Inadequate methods include using an open random allocation schedule (e.g., a list of random numbers); assignment envelopes used without appropriate safeguards (e.g., if envelopes were unsealed or nonopaque or not sequentially numbered); alternation or rotation; date of birth; case record number; or any other explicitly unconcealed procedure.*
	NR	There is insufficient information provided about allocation to study groups
	--	There is direct evidence that at the time of recruitment it was possible for the research personnel and subjects to know what study group subjects were allocated to, or it is likely that they could have broken the blinding of allocation before recruitment was complete and irrevocable
5. Were the research personnel and human subjects blinded to the study group during the study	++	There is direct evidence that the subjects and research personnel were adequately blinded to study group, **AND** it is unlikely that they could have broken the blinding during the study. Methods used to ensure blinding include central allocation; sequentially numbered drug containers of identical appearance; sequentially numbered, opaque, sealed envelopes; or equivalent methods
	+	There is indirect evidence that the research personnel and subjects were adequately blinded to study group, **AND** it is unlikely that they could have broken the blinding during the study **OR** it is deemed that lack of adequate blinding during the study would not appreciably bias results (this would depend on the outcome).
	-	There is indirect evidence that it was possible for research personnel or subjects to infer the study group **NOTE:** *Inadequate methods include using an open random allocation schedule (e.g., a list of random numbers), assignment envelopes used without appropriate safeguards, alternation or rotation; date of birth; case record number; or any other explicitly unconcealed procedure.*
	NR	There is insufficient information provided about blinding to study group during the study
	--	There is direct evidence for lack of adequate blinding of the study group including no blinding or incomplete blinding of research personnel and subjects. For some treatments, such as behavioural interventions, allocation to study groups cannot be concealed
6. Were outcome data completely reported without attrition or exclusion from analysis?	++	There is direct evidence that there was no loss of subjects during the study and outcome data were complete, **OR** loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study or analyses. Review authors should be confident that the participants included in the analysis are exactly those who were randomized into the trial. Acceptable handling of subject attrition includes: very little missing outcome data (e.g. < 10% in each group); reasons for missing subjects unlikely to be related to outcome; missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups, **OR**

		analyses (such as intention-to-treat analysis) in which missing data have been imputed using appropriate methods (insuring that the characteristics of subjects lost to follow up or with unavailable records are described in identical way and are not significantly different from those of the study participants). **NOTE:** *Participants randomized but subsequently found not to be eligible need not always be considered as having missing outcome data.*
	+	There is indirect evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study, **OR** it is deemed that the proportion lost to follow-up would not appreciably bias results (e.g. < 20% in each group). This would include reports of no statistical differences in characteristics of subjects lost to follow up or with unavailable records from those of the study participants. Generally, the higher the ratio of participants with missing data to participants with events, the greater potential there is for bias. For studies with a long duration of follow-up, some withdrawals for such reasons are inevitable.
	-	There is indirect evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large (e.g. > 20% in each group) and not adequately addressed
	NR	There is insufficient information provided about numbers of subjects lost to follow-up
	--	There is direct evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed. Unacceptable handling of subject attrition includes: reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across study groups; or potentially inappropriate application of imputation.
7. Can we be confident in the exposure characterisation? **Key question**	++	There is direct evidence that the exposure was adequately assessed, i.e. the sugar content of the intervention (and control) foods and/or beverages was measured during the study by e.g. food analysis **AND** there is direct evidence that the exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups (e.g., administration of study foods or diets was supervised; compliance was assessed using biomarkers of intake such as 24-h urinary excretion of sucrose or fructose).
	+	There is indirect evidence that the exposure was adequately assessed, i.e. the sugar content of the intervention (and control) foods and/or beverages was not measured but rather e.g. calculated from food composition tables, provided by the food manufacturer, calculated from the ingredients list; **AND** there is indirect evidence that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups (e.g. administration of study foods or diets was not supervised **but** study products were provided by the investigators and compliance was assessed using food records, return of unconsumed foods, or a similar method).
	-	There is indirect evidence that the exposure was assessed using poorly validated methods (e.g. study products or diets were not provided by the investigators and compliance was not checked; unclear portion sizes and sugar content of the foods consumed by the subjects)
	NR	There is insufficient information provided about the validity of the exposure assessment method (e.g. no information about how the sugars content of foods and beverages was estimated, no information on compliance), but no evidence for concern
	--	There is direct evidence that the exposure was assessed using poorly validated methods

		OR There is direct evidence of poor compliance with the intervention
8. Can we be confident in the outcome assessment? **Key question**	++	There is direct evidence that the outcome was assessed using well-established methods (e.g., the "gold standard"). Such methods will depend on the outcome, but may include objectively measured with diagnostic methods, measured by trained investigators. For dental caries, such methods include clinical examination by a trained dentist, calibration of investigators (if multiple outcome assessors) and confirmation of diagnosis by radiography **AND** subjects had been followed for the same length of time in all study groups, **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes.
	+	There is indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard). Such methods will depend on the outcome, but may include proxy reporting of outcomes, mining data collected for other purposes. For dental caries, such methods include clinical examination by a trained dentist and calibration of investigators (if multiple outcome assessors), without confirmation of diagnosis by radiography **AND** subjects had been followed for the same length of time in all study groups, **OR** it is deemed that the outcome assessment methods used would not appreciably bias results (e.g. when there is no information about the method but standard measurements are most likely, e.g. blood lipids, body weight in a research setting), **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures.
	-	There is indirect evidence that the outcome assessment method is an insensitive instrument (e.g., a questionnaire used to assess outcomes with no information on validation), **OR** the length of follow up differed by study group, **OR** there is indirect evidence that it was possible for outcome assessors (including study subjects if outcomes were self-reported) to infer the study group prior to reporting outcomes **AND** it is deemed that the outcome assessment methods used could appreciably bias results
	NR	There is insufficient information provided about blinding of outcome assessors **OR** there is no information about the outcome assessment method **AND** it is deemed that the outcome assessment method could have biased the results (e.g. for waist circumference, if outcome assessors are not blinded)
	--	There is direct evidence that the outcome assessment method is an insensitive instrument, **OR** the length of follow up differed by study group, **OR**

		there is direct evidence for lack of adequate blinding of outcome assessors (including study subjects if outcomes were self-reported), including no blinding or incomplete blinding **AND** it is deemed that the outcome assessment method could have biased the results
9. Were all measured outcomes reported? ***NOTE:*** *It is recognised that selective reporting is difficult to assess with confidence for most studies unless the study protocol is available. Selective reporting bias can be assessed by comparing the "methods" and "results" section of the paper, and by considering outcomes measured in the context of knowledge in the field. Selective reporting bias may be suspected if the study does not report outcomes in the results section that would have been expected based on the methods, or if a composite score is present without the individual component outcomes.*	++	There is direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
	+	There is indirect evidence that all the study's measured outcomes (primary and secondary) outlined in the methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, **OR** analyses that had not been planned in advance (i.e. retrospective unplanned subgroup analyses) are clearly indicated as such and it is deemed that the unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g. appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
	-	There is indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported **OR** there is indirect evidence that unplanned analyses were included that may appreciably bias results
	NR	There is insufficient information provided about selective outcome reporting
	--	There is direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results
10. Were there no other potential threats to internal validity (e.g. statistical methods were appropriate and researchers adhered to the study protocol)? ***NOTE:*** *Baseline characteristics should be appraised* ***only if*** *Q1 (randomisation) was rated with ++/+ and Q2 (allocation concealment) was rated with ++/+/NR*	++	There is direct evidence that variables, other than the exposure and outcome, did not differ between groups during the course of the intervention in a way that could bias results / For cross-over trials: there is direct evidence of no carry-over effects **AND** there is no evidence of differences in baseline characteristics between groups
	+	There is indirect evidence that variables, other than the exposure and outcome, did not differ between groups during the course of the intervention in a way that could bias results/ For cross-over trials: there is indirect evidence of no carry-over effects (e.g. presence of a sufficient washout period) **AND** there is no evidence of differences in baseline characteristics between groups **OR** there is evidence that reported variables differed between groups at baseline / For cross-over trials: no washout period **AND** It is deemed that these differences (or absence of washout for cross-over trials) would not appreciably bias results (no concern or adequately addressed by analysis)

	-	There is no information on variables, other than the exposure and outcome, which could bias the results would have differed between groups during the course of the intervention / For cross-over trials: no washout period **AND** there is indirect evidence that variables, other than the exposure and outcome, may have differed between groups during the course of the intervention in a way that could bias results/ For cross-over trials: indirect evidence of carry-over effects
	NR	There is no information about baseline characteristics by group (for parallel studies)
	--	There is evidence that variables, other than the exposure and outcome, differed between groups during the intervention / For cross-over trials: direct evidence of carry-over effects **AND** It is deemed that these differences appreciably biased results (there is concern e.g. not adequately addressed by analysis) **OR** there is evidence that reported variables differed between groups at baseline **AND i**t is deemed that these differences appreciably biased results (e.g. not adequately addressed by analysis)

3. Observational studies on metabolic diseases including pregnancy endpoints and dental caries

The risk of bias (RoB) in observational studies was assessed by type of exposure (nutrients and food sources separately) and per endpoint. This is because different elements were considered when evaluating the confidence in the exposure assessment for nutrients (e.g. total sugars, fructose, etc) (question 4a) versus food sources (e.g. SSBs, fruit juices) (Question 4b).

Key confounders considered for each endpoint are summarised in Table I2. Customised appraisal forms are in Table I3.

Table I2. Key confounders by endpoint

Confounders	**Endpoints**						
	Obesity /abdominal obesity	**T2DM**	**HTN**	**CVDs**	**Gout**	**Dental caries**	**GDM**
Age	x	x	x	x	x	x	x
Sex	x	x	x	x	x	x	
Socioeconomic status	x	x				x	x
Ethnicity (where relevant)	x	x	x	x			x
Energy intake	x	x	x	x	x		x
Alcohol intake				x	x		
Smoking			x	x			
Physical activity	x	x	x	x	x		x
BMI or WC as appropriate	x	x					x
Markers of obesity [(a)]		x	x	x	x		x
Markers of "diet quality" [(b)]	x	x	x	x			x
Intake of meat, seafood, purine rich vegetables					x		
Family history of diabetes		x					x
History of hypercholesterolemia/treatment				x			
History of hypertension/treatment				x	x		
History of chronic renal failure					x		
Fluoride use/tooth brushing						x	

(a) For studies on the relationship between the intake of sugar-sweetened beverages or fruit juice, body weight/BMI may be on the causal pathway between the exposure and the outcome. body weight/BMI are typically introduced at the end of the adjustment strategy to test for this hypothesis. The issue of over-adjustment has been addressed under Question 7

(b) to account for the potential confounding effect of the rest of the diet (e.g. dietary patterns associated with the consumption of sugar-sweetened beverages or fruit juice)

For birthweight-related endpoints (four studies), a list of key confounders was not defined a priori because the endpoints of interest were at the two extremes of the same variable (birthweight) and was difficult to anticipate the factors that could confound reported associations outside the context of the specific analytical strategy, depending of the objectives of the study. Risk of bias was judged by two reviewers and discussed within the working group. The handling of potential confounding for the Cadmen study (Lenders et al., 1997) and MoBa study (Grundt et al., 2017) was considered adequate (+, "probably low risk of bias"). In contrast, the HSS-USA study (Crume et al., 2016) was found to be at "probably high risk of bias" (-) due to inadequate adjustment for energy intake and the GeliS study (Günther et al., 2019) at "definitely high risk of bias" (- -) due to the lack of adjustment for gestational diabetes mellitus, pregnancy hypertension or preterm delivery at baseline.

Table I3. Customised form for the appraisal of human observational studies

[++: definitely low risk of bias, +: probably low risk of bias, -/NR: probably high risk of bias, --: definitely high risk of bias].

Question	Rating	Explanation for expert judgement
		Note: The scope of this question is limited to the appraisal of the risk of residual confounding (lack of adjustment for potential confounders). Issues related to potential over-adjustment are addressed under question 7 (other threats to internal validity). If a study adequately addressed all potential confounders identified, a + is given by default due to the limitations of the measurement methods which are typically used for measuring physical activity and energy intake; it may be upgraded to ++ if robust methods were applied to measure these variables
4. Did the study design or analysis account for important confounding and modifying variables? **Key question**	++	There is direct evidence that appropriate adjustments or explicit considerations were made for primary covariates and confounders in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, AND there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, AND there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for.
	+	There is indirect evidence that appropriate adjustments were made, OR it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results. AND there is evidence (direct or indirect) that primary covariates and confounders were assessed using valid and reliable measurements, OR it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), AND there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, OR it is deemed that co-exposures present would not appreciably bias results. *Note: This includes insufficient information provided on co-exposures in general population studies.*

	-/NR	There is indirect evidence that the distribution of primary covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, OR there is insufficient information provided about the distribution of known confounders (record "NR" as basis for answer), OR there is indirect evidence that primary covariates and confounders were assessed using measurements of unknown validity, OR there is insufficient information provided about the measurement techniques used to assess primary covariates and confounders (record "NR" as basis for answer), OR there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for, OR there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).
	--	There is direct evidence that the distribution of primary covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, OR there is direct evidence that primary covariates and confounders were assessed using non valid measurements, OR there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.
		Note: When attrition rate is not discussed/reported, attrition can be assumed to be low in view of the ascertainment method used for the identification of cases (e.g. national registry of diseases).
6. Were outcome data complete without attrition or exclusion from analysis?	**++**	There is direct evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when subjects were removed from a study. Acceptable handling of subject attrition includes: very little missing outcome data; reasons for missing subjects unlikely to be related to outcome (for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups (i.e. unlikely to be related to exposure), OR missing data have been imputed using appropriate methods and characteristics of subjects lost to follow up or with unavailable records are described in identical way and are not significantly different from those of the study participants.
	+	There is indirect evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when subjects were removed from a study, OR it is deemed that the proportion lost to follow-up would not appreciably bias results. This would include similarity between the characteristics of subjects lost to follow-up and study participants. *Note: Generally, the higher the ratio of participants with missing data to participants with events, the greater potential there is for bias. For studies with a long duration of follow-up, some withdrawals for such reasons are inevitable.*
	-/NR	There is indirect evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed,

		OR there is insufficient information provided about numbers of subjects lost to follow-up (record "NR" as basis for answer).
	--	There is direct evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed. Unacceptable handling of subject attrition includes: reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across study groups (i.e. likely to be related to the exposure); or potentially inappropriate application of imputation.
	++	Not applicable
7. a. Can we be confident in the exposure characterization? **Nutrient category** (e.g. total sugar, added sugar, total glucose) **Key question**	+	Method for classifying subjects according to their nutrient intake (at measurement timepoint): Semi-quantitative food diaries or 24h DR considered to have adequately captured within person variations OR Valid and reliable SFFQ or diet history AND Method to correct for measurement errors is applied (e.g. calibration, standardisation for energy (e.g. energy density, residual method) before categorisation & exclusion of implausible energy intake) *Note: Method for classifying subjects considering within-person changes in dietary patterns over "medium-/long-term": not of concern for nutrient categories in adults* **CHILDREN ONLY:** Method for classifying subjects considering within-person changes in dietary patterns over "medium-/long- term": Repeated measurements during follow up incorporated in the analysis or other method to address within-person changes in dietary patterns
	-/NR	Starting point for all methods
	--	Direct evidence of low validity *Note: In case of low (Spearman or Pearson) correlation coefficients, it is proposed not to downgrade the study by default but to check how authors commented on those and how they took them into consideration.*
	++	Not applicable
7.b. Can we be confident in the exposure characterization? **Food category** (e.g. SSSD, Fruit juices) **Key question**	+	Method for classifying subjects according to their food intake (at measurement timepoint): Semi-quantitative food diaries or 24h DR considered to have adequately captured within person variations OR Valid and reliable SFFQ or diet history **Note:** Insight in the validity of a SFFQ can be obtained from the formulation of the specific questions used to capture the exposure of interest (e.g. how specific (e.g. SSFD vs TFJ vs 100%FJ), unambiguous (diet soda vs SSSD)) AND Method for classifying subjects considering within-person changes in dietary patterns over "medium-/long- term": Repeated measurements during follow up incorporated in the analysis or other method to address within-person changes in dietary patterns
	-/NR	Starting point for all methods

	--	Direct evidence of low validity
8. Can we be confident in the outcome assessment? **Key question**	++	Use of objective and systematic method for diagnosis **Note:** **T2DM:** e.g. participants screened based on biochemical markers (fasting glucose, glucose at 2 hours during OGTT, glycated haemoglobin for diagnostic purposes) **Obesity/Abdominal obesity**: e.g. body weight & height, waist circumference measured by trained personnel **CVD events**: e.g. clinical diagnosis based on valid biomarkers (medical records) **Hypertension:** e.g. diagnosed based on measures of BP by trained personnel and standard SBP/DBP cut-offs for hypertension **Dental caries:** e.g. clinical examination by a trained dentist, calibration of investigators (if multiple outcome assessors) and confirmation of diagnosis by radiography
	+	Use of registry(ies) or self-reported AND Measures applied to minimize false negative (e.g. combination of methods) and false positive (e.g. cases verified based on objective information) **Note**: For fatal events such as fatal CHD, the use of national registry(ies) without further measures to minimize false positives may be considered sufficiently reliable (e.g. cause of death coded according to international standards); for non-fatal cases, the use of hospital discharge registries without further measures to minimize false positives may be considered sufficiently reliable. For dental caries: clinical examination by a trained dentist and calibration of investigators (if multiple outcome assessors), without confirmation of diagnosis by radiography
	-/NR	Use of registry(ies) or self-reported AND No measures applied to minimize false negative or false positive (including unspecific diagnostic criteria) OR There is insufficient information provided about the outcome measurement method
	--	Direct evidence that different diagnostic tools have been used differentially across exposure groups
10. Were there no other potential threats to internal validity?		**Note:** This includes: a) the appraisal of potential over-adjustment, i.e. adjustment for variables which are not confounders b) the appraisal of bias due to selective reporting
	++	There is no evidence for other threats to internal validity (e.g. regarding overadjustment or selective reporting)
	+	There is direct/indirect evidence for other threats to internal validity (e.g. regarding overadjustment or selective reporting) but it is deemed that the issue identified would not appreciably bias results
	-/NR	There is indirect evidence for other threats to internal validity (e.g. regarding overadjustment or selective reporting) that could appreciably bias results OR There is insufficient information to appraise this question
	--	There is direct evidence for other threats to internal validity (e.g. regarding overadjustment or selective reporting) that would appreciably bias results

Glossary, abbreviations, and acronyms

100% FJ	100% FJ, with no added sugar
24h DR	24 hour dietary record
BMI	Body mass index
BP	Blood pressure
CHD	Coronary heart disease
CVDs	Cardiovascular diseases
DBP	Diastolic blood pressure
HDL	High density lipoprotein
HTN	Hypertension
IVGTT	Intravenous glucose tolerance test
LDL	Low density lipoprotein
NR	Not reported
NTP	National Toxicology Program
OGTT	Oral glucose tolerance test
RoB	Risk of bias
SBP	Systolic blood pressure
SFFQ	Semi-quantitative food frequency questionnaire
SSBs	Sugar-sweetened beverages
SSFD	Sugar-sweetened fruit drink
SSSD	Sugar-sweetened soft drink
T2DM	Type 2 diabetes mellitus
TFJ	Total fruit juice
WC	Waist circumference

References

Crume TL, Brinton JT, Shapiro A, Kaar J, Glueck DH, Siega-Riz AM and Dabelea D, 2016. Maternal dietary intake during pregnancy and offspring body composition: The Healthy Start Study. Am J Obstet Gynecol, 215:609.e601-609.e608. doi: 10.1016/j.ajog.2016.06.035. Epub 2016 Jun 28.

Grundt JH, Eide GE, Brantsaeter AL, Haugen M and Markestad T, 2017. Is consumption of sugar-sweetened soft drinks during pregnancy associated with birth weight? Matern Child Nutr, 13. doi: 10.1111/mcn.12405. Epub 2016 Dec 7.

Günther J, Hoffmann J, Spies M, Meyer D, Kunath J, Stecher L, Rosenfeld E, Kick L, Rauh K and Hauner H, 2019. Associations between the Prenatal Diet and Neonatal Outcomes-A Secondary Analysis of the Cluster-Randomised GeliS Trial. Nutrients, 11. doi: 10.3390/nu11081889.

Lenders CM, Hediger ML, Scholl TO, Khoo CS, Slap GB and Stallings VA, 1997. Gestational age and infant size at birth are associated with dietary sugar intake among pregnant adolescents. J Nutr, 127:1113-1117. doi: 10.1093/jn/127.6.1113.